DOCUMENT RESUME

ED 265 190 TM 860 042

AUTHOR Fuchs, Douglas; And Others
TITLE Norm-Referenced Tests: Are They Valid for Use with Handicapped Children?
PUB DATE [85]
NOTE 33p.; Portions of this paper were presented at the Annual Meeting of the and at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 16-20, 1986).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; *Aptitude Tests; *Disabilities; Elementary Secondary Education; Interrater Reliability; *Norm Referenced Tests; Special Education; Standardized Tests; *Test Construction; *Test Interpretation; Test Manuals; Test Reliability; Test Validity

ABSTRACT

The present investigation represents a systematic effort to determine whether handicapped children have been included in the development of test norms, items, and indices of reliability and validity. It analysed up-to-date user manuals and technical supplements of 27 well known and widely used aptitude and achievement tests. Study procedure involved four basic steps: (1) selecting tests; (2) contacting publishers for current and complete technical information; (3) analyzing the technical data; and (4) determining interrater agreement. A matrix was constructed to include the tests and six psychometric characteristics, yielding 162 cells. Each cell was filled in or left blank, according to the information pertaining to the inclusion of the handicapped. Sixty-three cells were blank; 84 contained question marks, indicating widespread failure among test developers to clarify the population on which they developed instruments. Results indicated most test developers and publishers provide scant data on the appropriateness of their tests for use with handicapped children. Appendices include a four-page bibliography, a table showing the titles, authors, publishers and copyright dates of the norm-referenced tests, and a table showing handicapped children's participation in the development of the tests. (LMO)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

# Norm-Referenced Tests: Are They Valid for Use with Handicapped Children?

ED265190

Douglas Fuchs and Lynn S. Fuchs

Peabody College of Vanderbilt University

Suanne Benowitz

Clark University

Kenneth Barringer

Newton Academy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Fuchs, D.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Portions of this paper were presented at the 1985 annual meeting of the Council for Exceptional Children in Anaheim and the 1986 annual meeting of the American Educational Research Association in San Francisco.

Requests for reprints should be sent to: Douglas Fuchs, Box 328, Department of Special Education, Peabody College, Vanderbilt University, Nashville, TN 37203.

Running head: Norm-Referenced Tests

TM 860042

Abstract

Although norm-referenced tests have been a mainstay in special education placement and program evaluation activity, there never has been a systematic effort to determine whether handicapped children have been included in the development of the tests' norms, items, and indices of reliability and validity. The present investigation explored this question by analyzing up-to-date user manuals and technical supplements of 27 aptitude and achievement tests reviewed by Salvia and Ysseldyke (1978). Results indicated most test developers and publishers provide scant data on the appropriateness of their tests for use with handicapped children. Implications for test users, developers, and publishers are discussed.

# Norm-Referenced Tests: Are They Valid for Use with Handicapped Children?

Recently there has been increased use of criterion-referenced tests. Nevertheless, norm-referenced measures, including IQ, achievement, and screening instruments, continue to be employed widely in special education, providing data for diagnostic, classification, program placement, and evaluation decisions. Despite or perhaps because of their long-term use, norm-referenced tests have been the focus of much criticism.

Special education teachers and administrators, parents and officers of the court, professional organizations responsible for establishing testing standards, and educational researchers have expressed a variety of concerns. Critics contend that (a) test content discriminates against certain subgroups of the population (e.g., Cole & Bruner, 1972; Gould, 1981; McClelland, 1973; Ogbu, 1978), (b) the technical characteristics of many tests are inadequate (e.g., Ysseldyke & Shinn, 1981), (c) typical testing procedures are biased against low SES examinees (Fuchs & Fuchs, in press), (d) many tests both fail to explore pupils' cognitive processes ( Gallas & Sigel, 1979) and to guide instruction (e.g., Hunt, 1975; Jenkins & Pany, 1978), (e) tests facilitate the use of pejorative labels (e.g., Reynolds & Balow, 1972; Scriven 1975), which, in turn, negatively affect teacher judgments (e.g., Foster, Ysseldyke, & Reese, 1975; Palmer, 1983), and (f) examiners can be incompetent in selecting appropriate tests (e.g., Keogh, Kukic, Becker, McLoughlin, & Kukic, 1975), inaccurate in administering and scoring tests (e.g., Bennett, 1983), and imprudent in failing to base instructional recommendations

and decisions on available assessment data (e.g., Ysseldyke & Algozzine, 1982).

Another important issue, one which has been discussed much less frequently, is whether test constructors employ handicapped children in the development of test norms, items, and indices of reliability and validity. This concern draws much of its legitimacy and importance from recent federal legislation that demands an end to the discrimination of the handicapped and encourages their participation in the mainstream of American life.

Both PL94-142 and Section 504 of the Rehabilitation Act of 1973 recognize that testing may discriminate against the handicapped. Such recognition is implicit in Section 504, which establishes placement procedures to ensure that "children are not misclassified, unnecessarily labeled as handicapped, or incorrectly placed because of inappropriate selection, administration, or interpretation of evaluation materials" (42 Fed. Reg. 22691, cited in Sherman & Robinson, 1982). To eliminate the use of placement procedures that discriminate against the handicapped, Section 504 requires that tests be validated for the specific purpose for which they are used. In other words, they should be selected and administered so that, when given to handicapped students, results accurately reflect pupils' aptitude, achievement, or whatever other factors the tests purport to measure rather than reflecting the children's disabilities (except, of course, when the specific disabilities are the factors that the tests claim to measure). In short, before tests may be considered nondiscriminatory, or unbiased estimates of handicapped children's potential, they should measure the same abilities and skills among handicapped and nonhandicapped groups.

Contributing to the seriousness of this concern is the apparent complexity involved in the development of tests that are valid for use with both handicapped and nonhandicapped groups. As described by the National Research Council's Panel on Testing Handicapped People (Sherman & Robinson, 1982), major modifications in the medium and method of test administration, as well as changes in test content, often are necessary before many handicapped children may be evaluated in a valid manner. For example, visually impaired children collectively may require a variety of modifications in the medium in which a test is administered: It may be necessary to produce test booklets in large print, high quality regular print, or braille, to tape record the test, or to read it aloud to the child. The method used to record a response also may require alteration: Visually impaired examinees who cannot record their answers may need the assistance of a tape recorder, typewriter, braille writer, or person who writes or marks the answers.

Likewise, changes in test content may be required. Items may be unnecessarily difficult (i.e. discriminatory) for a visually impaired child if they measure knowledge, skills, or concepts learned primarily through vision or if the items use visual stimuli to measure knowledge acquired through other senses. Further complicating the formulation of valid tests for the handicapped is that the modifications developed for appropriate evaluation of visually impaired children may be qualitatively different from modifications necessary for the accurate testing of hearing impaired pupils; that is, children with contrasting handicapping conditions may require substantively different changes in the medium and method of test administration and test content.

Given the importance and apparent difficulty of developing normative tests that measure the same abilities and skills among handicapped and nonhandicapped children, the obvious question is, "Do they?" An important step toward answering this question is to explore test developers' user manuals and technical supplements to determine whether, and if so to what extent, handicapped children participated in the development of widely used normative tests. Although there have been occasional assertions that handicapped children have been excluded from most normative tests' development (Bennett, 1983; Jones, 1973; Sherman & Robinson, 1982), we know of no systematic empirical attempt to determine the degree of participation of handicapped children in the creation of well known and widely used tests' norms, items, and indices of reliability and validity. This was the primary purpose of the present study.

## Method

### Procedure

The study's procedure included four basic steps: selecting tests, contacting publishers for current and complete technical information, analyzing the technical data, and determining interrater agreement.

Selecting tests. Table 1 lists the titles, authors, publishers, and copyright dates of the 27 tests included in the analysis. These represent all but one of the tests reviewed by Salvia and Ysseldyke (1978) in the following categories and sub-categories: (a) norm-referenced achievement tests, both screening and diagnostic, and (b) individually administered IQ tests, including global measures, picture vocabulary tests yielding IQ scores, and IQ tests designed for specific populations. These tests were selected, and presumably were

reviewed by Salvia and Ysseldyke, because they are well known and widely used in special education.

---

Insert Table 1 about here

---

Supporting this contention are data from a survey conducted by Goh, Teslow, and Fuller (1981) of a geographically representative sample of members of the National Association of School Psychologists. Respondents were asked to name the eight tests they used most often in each of a number of assessment areas. Among the tests reviewed by Salvia and Ysseldyke, and included in this study, 14 measures were mentioned by the psychologists: 6, 4, and 4 in the areas of Intelligence, Achievement-General, and Achievement-Specific, respectvely. In a similar investigation, Mardell-Czudnowski (cited in LaGrow & Prochnow-LaGrow, 1982) surveyed one-fifth of all licensed full-time school psychologists employed in Illinois in 1980. The survey asked respondents to identify the most frequently used tests in special education. The five aptitude or achievement tests named most often are among the 27 tests included in this study.

Contacting publishers. The publishers of each test were contacted for two reasons: to determine whether we had the most recent user manual and to explore whether additional technical information was available in published supplements, bulletins, or handbooks. If we had out-dated or incomplete test information, we ordered and obtained the most recent user manual and all published supplementary technical information.

Analyzing technical data. The user manual and /or technical supplement of every test was analyzed in terms of norms, item

development, internal and test-retest reliability, and concurrent and predictive validity. A matrix was constructed to include the 27 tests (rows) and 6 psychometric characteristics (columns), yielding 162 cells. For each test and every psychometric property, we inquired about handicapped children's inclusion. When no data were presented for a psychometric characteristic of a given test, the corresponding cell in the matrix was left blank. If it was stated that handicapped children participated in the development of a psychometric property, but no percentage of involvement was specified, a "Yes" was recorded. When the percentage of handicapped children's involvement was made clear, the specific percentage was entered. Finally, if the test manual and technical supplement were unclear about handicapped children's participation, a question mark (?) was recorded.

Determining interrater agreement. Interrater agreement was determined using the formula suggested by Coulter (cited in Thompson, White, & Morgan, 1982): Percentage of agreement = agreements between raters A and B/agreements + disagreements between raters A and B + omissions by rater A + omissions by rater B. Calculated on 100% of the data entered in the matrix, interrater agreement was .84. Disagreements were resolved through discussion between the two raters.

## Results

Table 2 displays the matrix described above. Two distinctive characteristics of the matrix are its numerous question marks and blank spaces. Each of these features represents important and related facts about the 27 popular tests.

---

Insert Table 2 about here

---

Question Marks

The question marks are of two types: those with and without footnotes. As discussed above, those without footnotes signify the failure of test developers to clarify whether they included handicapped children in the standardization populations, development of test items, and establishment of indices of reliability and validity. Many of the question marks with footnotes indicate the existence of technical data generated on handicapped children, but this information was not produced by the test developers. Instead, these data typically were produced by researchers uninvolved in test development. These researchers frequently employed small and homogeneous groups of children, who were not part of any normative population, and sometimes administered the tests in ways different from the procedures used by the test developers. Because of the size and nature of these groups and the sometimes incomparable ways in which test developers and researchers administered the tests, we believe the technical data referenced by the footnotes may be inadequate and potentially misleading. Additionally, since some of this information is not readily available to most test users, we believe it does not always represent a useful data base. Thus, for the majority of footnoted question marks in the matrix, we see no reason to distinguish them from the question marks without footnotes.

Regarding the question marks in an undifferentiated manner indicates that they constitute 52% of the matrix (or 84 of 162 cells). Although no test developer explicitly excluded the handicapped from

the norms, item development, or establishment of indices of reliability and validity, only 5 of 27 tests provide specific percentages of handicapped children's participation in some aspect of test development. These five tests account for a mere 12 of 162 cells, or 7% of the matrix, displaying specific percentages of handicapped children's involvement. In three additional cells (2% of the matrix), item development of the McCarthy Scales of Children's Abilities and Slosson Intelligence Test and norms of the Silent Reading Diagnostic Test, a "yes" is displayed, indicating handicapped children were involved but at an unspecified level. This widespread vagueness extends even to tests specifically designed for or widely used with special populations. Among IQ tests for special populations shown in Table 2, only the Blind Learning Aptitude Test and Nebraska Test of Learning Aptitude provide specific percentages of handicapped children's participation in test development.

Blank Spaces

As with the question marks, some blank spaces in Table 2 are footnoted and some are not. Those without footnotes mark instances in which test constructors do not provide users with any technical information at all. Many of the blank spaces with footnotes indicate technical data exist, but were not produced by the test developers. We believe much of these data may be inadequate and misleading for many of the same reasons discussed with respect to the footnoted question marks.

Viewing the blanks without regard to footnotes indicates that many of the developers of the 27 popular tests offer scant technical data: The Stanford-Binet fails to report indices of internal or test-retest reliability or of concurrent or predictive validity; the

Leiter presents no data on the nature of its norms, item development, internal and test-retest reliability, or concurrent and predictive validity; the Peabody Picture Vocabulary Test-Revised, Metropolitan and Stanford Achievement Tests, Wide Range Achievement Test-Revised, Gray Oral Reading Test, Silent Reading Diagnostic Test, and Woodcock Reading Mastery Tests have neither concurrent nor predictive validity information; and the Quick Test does not provide pertinent reliability data. For a psychometric property such as predictive validity, only 3 of the 27 tests possess pertinent data.

Across all 162 cells of the matrix, 63 cells (39%) are blank. Combining these 63 cells with the 84 additional ones containing question marks, reveals that, for 147 of 162 cells (91% of the matrix), test developers provide no information on the nature and extent of handicapped children's involvement in test development.

## Discussion

Findings indicate that, among 27 widely used normative tests, numerous test constructors provide no evidence to indicate their tests are valid for use with handicapped children. A majority of cells in the matrix displayed in Table 2 contain question marks, indicating widespread failure among test developers to clarify the population on which they developed their instruments. It is possible test constructors employed handicapped children (without identifying them as handicapped) more frequently than our findings indicate. However, until many of them describe more completely the children on whom they have developed their tests, they and their publishers are in explicit violation of widely recognized professional standards for test development.

The Standards for Educational and Psychological Testing (Standards; American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985) call for clear and comprehensive descriptions of those employed in the development of test norms (see Standard 4.3, p. 33), items (see Standard 3.12, p. 28), and indices of reliability (see Standard 2.2, p. 20) and validity (see Standard 1.5, p. 14). If, in fact, test constructors have not validated their instruments for use with handicapped people, they ". . . should issue cautionary statements in manuals and elsewhere regarding confidence in interpretations . . ." based on these tests (Standard 14.2, p. 79). Moreover, if such ". . . a test is likely to be used incorrectly for certain kinds of decisions, specific warnings against such use should be given" (Standard 1.2, p. 13). In our analysis of the manuals and technical supplements of the 27 popular tests, we rarely found admonishing qualifications in the information describing the tests' intended uses.

Underscoring the need for test developers to provide data on the use of their instruments with handicapped people is empirical evidence indicating typical test procedures discriminate against certain handicapped children's performance. In a series of studies, Fuchs and associates have shown that speech- and language-impaired children do not perform optimally when tested by an examiner with whom they are unfamiliar (see Fuchs, Featherstone, Garwick, & Fuchs, 1984; Fuchs, Fuchs, Dailey, & Power, 1985; Fuchs, Fuchs, Garwick, & Featherstone, 1983). Most recently, Fuchs, Fuchs, Power, and Dailey (1985) demonstrated that, although speech- and language-handicapped children perform significantly less well with unfamiliar than with familiar examiners, nonhandicapped children perform equally well in

both conditions.

These findings have important practical implications because of two apparent features of many normative tests. First, an average normative score probably reflects the mean performance of nonhandicapped children. Second, during the development of the norms, tests typically are administered by unfamiliar examiners (Fuchs, 1981). The research of Fuchs, Fuchs, Power, and Dailey (1985) indicates the use of unfamiliar examiners in the norming process does not depress the performance of the nonhandicapped standardization population. However, when speech- and language-impaired children are assessed by unfamiliar examiners, their suboptimal performance is compared to the more optimal performance of the nonhandicapped normative group. In such comparisons, examiner unfamiliarity is a source of systematic error or bias that invalidates test-related diagnostic, classification, and program placement decisions.

Documenting the important fact that many test developers do not provide validity information on the use of their instruments with handicapped children does not absolve test users from the responsibility to employ appropriate tests and procedures with this group. Rather, according to the Standards, the absence of this information increases users' obligations to test handicapped pupils in a manner that does not discriminate selectively against them. On page 42, Standard 6.3 states, "When a test is to be used for a purpose for which it has not been previously validated, or for which there is no supported claim for validity, the user is responsible for providing evidence of validity" (italics added). Because the Standards often are a guide in litigation, special education teachers, school psychologists, speech clinicians, and other test users should

recognize the potential liability in administering tests without pertinent validity information. Such recognition, however, may not be widespread. Evidence suggests that at least some examiners tend "to select tools in a routinized manner without consideration for the purposes of assessment, an instrument's technical adequacy for those purposes, or even the basic descriptive information contained in instrument manuals" (Bennett, 1983).

As mentioned at the outset, validating tests for the handicapped is a complex process. For test constructors and publishers this complexity no doubt represents additional expenditures of effort and money (see Sherman & Robinson, 1982). Whereas some "bottom line" test developers and publishers may regard validating their tests on the handicapped as a project they cannot afford to do, we view such an effort as something they "cannot afford not to do." Tests without validation data on the handicapped simply cannot and should not be used with this group for purposes of determining diagnoses, classifications, program placements, and evaluations of academic progress.

This is not only our view. It is the position expressed in PL94-142 and Section 504 of the Rehabilitation Act of 1973; it runs as a motif through the Standards; and, we believe, it represents the sentiments of many handicapped children and their families, teachers, and testers. Fortunately, some private institutions have made important contributions toward validating some tests for the handicapped. The American Printing House for the Blind, for example, produces and distributes large print and braille versions of many norm-referenced tests (W. Duckworth, personal communication, October 25, 1985). With respect to the 1982 edition of the Stanford

Achievement Test, Gallaudet College has modified testing procedures and established norms specifically for hearing-impaired children (see Allen, White, & Karchmer, 1983). Additionally, it is encouraging to note that at least some test developers and publishers have attempted to validate their tests for use with the handicapped (see Table 2), and that others, such as Educational Testing Service, are in the process of doing so (see Bennett & Ragosta, 1984).

References

Allen, T.E., White, C.S., & Karchmer, M.A. (1983). Issues in the development of a special edition for hearing-impaired students of the seventh edition of the Stanford Achievement Test. American Annals of the Deaf, 128, 34-39.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Bennett, R.E. (1983). Research and evaluation priorities for special education assessment. Exceptional Children, 50, 110-117.

Bennett, R.E., & Ragosta, M. (1984). A research context for studying admissions tests and handicapped populations (Report No. 1). Princeton, NJ: Educational Testing Service.

Cole, M., & Bruner, J.S. (1972). Preliminaries to a theory of cultural differences. In I. Gordon (Ed.), Seventy-first yearbook of the National Society for the Study of Education, part II -- Early Childhood Education. Chicago: University of Chicago.

Foster, G.G., Ysseldyke, J.E., & Reese, J. (1975). I never would have seen it if I hadn't believed it. Exceptional Children, 41, 469-474.

Fuchs, D. (1981, April). Differential responses of preschool language-handicapped children and familiar and unfamiliar testers as a function of task complexity, length of acquaintanceship, and sex of child. In V. Shipman (Chair), Client identification and issues of validity: The influence of situational variables on children's cognitive performance. Symposium presented at the

annual meeting of the American Educational Research Association, Los Angeles.

Fuchs, D., Featherstone, N.L., Garwick, D.R., & Fuchs, L.S. (1984). Effects of examiner familiarity and task characteristics on speech- and language-impaired children's test performance. Measurement and Evaluation in Guidance, 16, 198-204.

Fuchs, D., & Fuchs, L.S. (in press). Test procedure bias: A meta-analysis of examiner familiarity effects. Review of Educational Research.

Fuchs, D., Fuchs, L.S., Dailey, A.M., & Power, M.H. (1985). The effect of examiners' personal familiarity and professional experience on handicapped children's test performance. Journal of Educational Research, 78, 141-146.

Fuchs, D., Fuchs, L.S., Garwick, D.R., & Featherstone, N.L. (1983). Test performance of language-handicapped children with familiar and unfamiliar examiners. Journal of Psychology, 114, 37-46.

Fuchs, D., Fuchs, L.S., Power, M.H., & Dailey, A.M. (1985). Bias in the assessment of handicapped children. American Educational Research Journal, 22, 185-197.

Gallas, H.B., & Sigel, I.E. (1979). Cognitive-developmental assessment in children: Application of a cybernetic model. In M.N. Ozer (Ed.), A cybernetic approach to the assessment of children: Toward a more humane use of human beings. Boulder, CO: Westview Press.

Goh, D.S., Teslow, C.J., & Fuller, G.B. (1981). The practice of psychological assessment among school psychologists. Professional Psychology, 12, 696-706.

Gould, S.J. (1981). The mismeasure of man. New York: W.W. Norton.

Hunt, J.M. (1975). Psychological assessment in education and social class. In B.Z. Friedlander, G.M. Sterritt, & G.E. Kirk (Eds.), Exceptional infant (Vol. 3, pp. 545-574). New York: Brunner/Mazel.

Jenkins, J.R., & Pany, D.P. (1978). Standardized achievement tests: How useful for special education? Exceptional Children, 44, 448-453.

Jones, R.L. (1973). Accountability in special education: Some problems. Exceptional Children, 40, 631-642.

Keogh, B., Kukic, S., Becker, L., McLoughlin, R., & Kukic, M. (1975). School psychologists' services in special education programs. Journal of School Psychology, 13, 142-148.

LaGrow, S.J., & Prochnow-LaGrow, J.E. (1982). Technical adequacy of the most popular tests selected by responding school psychologists in Illinois. Psychology in the Schools, 19, 186-189.

McClelland, D.C. (1973). Testing for competence rather than for "intelligence." American Psychologist, 28, 1-14.

Ogbu, J.U. (1978). Minority education and caste. New York: Academic Press.

Palmer, D.J. (1983). An attributional perspective on labeling. Exceptional Children, 49, 423-429.

Reynolds, M.C., & Balow, B. (1972). Categories and variables in special education. Exceptional Children, 38, 357-366.

Salvia, J., & Ysseldyke, J.E. (1978). Assessment in special and remedial education. Boston: Houghton Mifflin.

Scriven, M. (1975). Some issues in the logic and ethics of mainstreaming. In J.C. Merwin (Ed.), Mainstreaming: Origins and

implications. Minneapolis: University of Minnesota.

Sherman, S.W., & Robinson, N.M. (1982). Ability testing of handicapped people: Dilemma for government, science, and the public. Washington, DC: National Academy Press.

Thompson, R.H., White, K.R., & Morgan, D.P. (1982). Teacher-student interaction patterns in classrooms with mainstreamed mildly handicapped students. Americ Educational Research Journal, 19, 220-236.

Ysseldyke, J.E., & Algozzine, B. (1982). Critical issues in special and remedial education. Boston: Houghton Mifflin.

Ysseldyke, J.E., & Shinn, M.R. (1981). Psychoeducational evaluation. In J.M. Kauffman & D.P. Hallahan (Eds.), Handbook of special education (pp. 418-440). Englewood Cliffs, NJ: Prentice Hall.

Table 1

Titles, Authors, Publishers, and Copyright Dates

of 27 Norm-Referenced Tests

| Type of test | Text name/abbreviation | Author(s) of manual or technical supplement/copyright | Publisher |
|---|---|---|---|
| General IQ tests | McCarthy Scales of Children's Abilities (MSCA) | McCarthy (1972) | Harcourt Brace Jovanovich |
| | Slosson Intelligence Test (SIT) | Slosson (1982)[1] | Slosson Educational Publications |
| | Stanford Binet Intelligence Scale (Binet) | Terman & Merrill (1973) | Riverside |
| | Wechsler Adult Ingelligence Scale-Revised (WAIS-R) | Wechsler (1981) | Harcourt Brace Jovanovich |
| | Wechsler Intelligence Scale for Children-Revised (WISC-R) | Wechsler (1974) | Harcourt Brace Jovanovich |
| | Wechsler Preschool and Primary Scale for Intelligence (WPPSI) | Wechsler (1967) | Harcourt Brace Jovanovich |
| IQ tests for special populations and Picture vocabulary tests | Blind Learning Aptitude Test (BLAT) | Newland (1969) | Harper and Row |
| | Columbia Mental Maturity Scale (CMMS) | Burgemeister, Blum, & Lorge (1972) | Harcourt Brace Jovanovich |
| | Leiter International Performance Scale (Leiter) | Levine (1983)[2] | Stoelting |

Table 1 (continued)

| Type of test | Test name/abbreviation | Author(s) of manual or technical supplement/copyright | Publisher |
|---|---|---|---|
| IQ tests for special populations and Picture vocabulary tests | Nebraska Test of Learning Aptitude (NTLA) | Hiskey (1966) | Marshall S. Hiskey |
| | Pictorial Test of Intelligence (PTI) | French (1964) | Houghton-Mifflin |
| | Peabody Picture Vocabulary Test-Revised (PPVT-R) | Dunn & Dunn (1981) | American Guidance Service |
| | Quick Test (QT) | Ammons & Ammons (1962)[3] | Psychological Test Specialists |
| Achievement screening tests | California Achievement Tests, Forms C and D (CAT) | CTB/McGraw-Hill (1979)[4] | CTB/McGraw-Hill |
| | Gates-MacGinitie Reading Tests (GMRT) | MacGinitie (1981)[5] | Riverside |
| | Metropolitan Achievement Tests Survey Battery (MAT) | Prescott, Balow, Hogan, & Farr (1978) | Harcourt Brace Jovanovich |
| | Peabody Individualized Achievement Test (PIAT) | Dunn & Markwardt (1970) | American Guidance Service |
| | Stanford Achievement Test (SAT) | Harcourt Jovanovich (1984)[6] | Brace Harcourt Brace Jovanovich |
| | Wide Range Achievement Test-Revised (WRAT-R) | Jastak & Wilkinson (1984) | Jastak Associates |

Table 1 (continued)

| Type of test | Test name/abbreviation | Author(s) of manual or technical supplement/copyright | Publisher |
|---|---|---|---|
| Achievement diagnostic tests | Diagnostic Reading Scales (DRS) | CTB/McGraw-Hill(1982)[7] | CTB/McGraw-Hill |
| | Durrell Analysis of Reading Difficulty (Durrell) | Durrell & Catterson (1980) | Harcourt Brace Jovanovich |
| | Gilmore Oral Reading Test (Gilmore) | Gilmore & Gilmore (1968) | Harcourt Brace Jovanovich |
| | Gray Oral Reading Test (Gray Oral) | Gray (1967) | Pro-Ed |
| | Reading Diagnostic Tests (RDT) | Gates, McKillop, & Horowitz (1981) | Teachers College |
| | Silent Reading Diagnostic Tests (SRDT) | Bond, Balow, & Hoyt (1970) | Meredith |
| | Stanford Diagnostic Reading Test/Red, Green, and Brown levels, (SDRT) | Karlson, Madden, & Gardner (1976) | Harcourt Brace Jovanovich |
| | Woodcock Reading Mastery Tests (WRMT) | Woodcock (1973) | American Guidance Service |

[1]Technical manual.

[2]Levine, M. N. (1983). Leiter international performance scale: A handbook. Chicago: Stoelting.

[3]Ammons, R. B., & Ammons, H. S. (1962). The Quick Test (QT): Provisional manual. Psychological Reports, 11, 111-161.

[4]CTB/McGraw Hill (1979). The California Achievement Tests: Technical bulletin No. 1. Monterey, CA: CTB/McGraw-Hill.

[5]MacGinitie, W. H. (1981). Technical summary. Chicago: Riverside.

[6]Harcourt Brace Jovanovich (n.d.). Technical review kit for the Stanford Achievement Series. New York: Harcourt Brace Jovanovich.

[7]CTB/McGraw-Hill. (1982). Diagnostic Reading Scales: Technical report. Monterey, CA: McGraw-Hill.

Table 2

Handicapped Children's Participation in the Technical Development of 27 Norm-Referenced Tests

| Type of test | Test name | Norms | Item development | Reliability: Internal | Reliability: Test-retest | Validity: Concurrent | Validity: Predictive |
|---|---|---|---|---|---|---|---|
| General IQ tests | MSCA | ?[1] | Yes[2] | ? | ? | ? | ? |
| | SIT | 24%[3] | Yes[4] | — | ?[5] | 24%[6] | — |
| | Binet | ? | ? | — | — | — | — |
| | WAIS-R | ?[7] | — | ? | ? | —[8] | —[8] |
| | WISC-R | ?[9] | — | ? | ? | —[10] | — |
| | WPPSI | ? | — | ? | ? | ? | — |
| IQ tests for special populations and Picture vocabulary tests | BLAT | 100%[11] | 100% | 100% | 100% | 100% | — |
| | CMMS | ? | ? | ? | ? | ?[12] | —[12] |
| | Leiter | — | — | —[13] | —[13] | —[13] | —[13] |
| | NTLA | 100%[14] | 100% | 100% | — | —[15] | — |
| | PTI | ? | ? | ? | ? | ?[16] | —[17] |
| | PPVT-R | ? | ? | ?[18] | ?[18] | — | — |
| | QT | ? | — | — | — | ?[19] | — |

Table 2 (continued)

| Type of test | Test name | Norms | Item development | Reliability: Internal | Reliability: Test-retest | Validity: Concurrent | Validity: Predictive |
|---|---|---|---|---|---|---|---|
| | CAT | ? | ? | ? | ? | ? | ___ |
| Achievement | GMRT | ? | ___ | ? | ? | ? | ___ |
| screening | MAT | ? | ? | ? | ___ | ___ | ___ |
| tests | PIAT | ? | ? | ___ | ? | ?[20] | ___ |
| | SAT | ? | ? | ? | ___ | ? | ___ |
| | WRAT-R | 2%[21] | ___ | ? | ? | ___[22] | ___ |
| | DRS | ?[23] | ? | ? | ? | ?[24] | ?[25] |
| | Durrell | ? | ? | ? | ___ | ___ | ?[26] |
| Achievement | Gilmore | ? | ? | ? | ___ | ?[27] | ___ |
| diagnostic | Gray Oral | ?[28] | ? | ___ | ___ | ___ | ___ |
| tests | RDT | ? | ___ | ___ | ? | ? | ___ |
| | SRDT | Yes[29] | ? | ? | ___ | ___ | ___ |
| | SDRT | ? | ? | ? | ___ | ? | ___ |
| | WRMT | ? | 3%[30] | ? | ? | ___ | ___ |

[1]Definition of "normal" seems to permit the possibility that mildly and moderately handicapped children were included (see page 16 of the manual).

[2]On page 1 it is stated that mentally retarded children were involved, but no percentages or numbers are presented.

[3]On page 21 of the manual it is reported that, among 1,109 standardization subjects, 266 were retarded with IQs less than 84 ($\bar{x}$ = 71.90, SD = 9.47).

[4]On page 3 it is stated that mentally retarded children and institutionalized adults were employed in item try-outs, but no percentage or number is given.

[5]The standardization sample was not divided into ability groups for purposes of determining test stability (see page 33).

[6]All 1,109 individuals in the normative group also completed the Binet. Also, concurrent validity studies employing the Binet, WISC, and WISC-R are reported on pages 55, 57, 58, and 60.

[7]Information on page 18 of the manual precludes the involvement of severely handicapped adults; it does not rule out the participation of mildly and moderately handicapped individuals.

[8]On pages 49-50 reference is made to validity studies conducted by others, which presumably did not involve the standardization sample.

[9]The definition of "normal" does not preclude the involvement of mildly and moderately handicapped children (see page 19 of the manual).

[10]Concurrent validity studies employing the WAIS, WPPSI, and Binet are reported on pages 48-51. Only the group of children tested on the Binet were part of the standardization population.

[11]All subjects were blind; there is no mention of the possibility of additional handicaps (see page 2 of the manual).

[12]On pages 42-47 of the manual there is an annotated bibliography of reliability and validity studies involving cerebral palsied, deaf, speech-impaired, mentally retarded, and neurologically-impaired children and either the 1954 or 1959 editions of the test.

[13]On pages 78-86 and 89-95 of the manual, reliability and validity studies are reported for handicapped individuals. None of these studies appeared to employ individuals constituting the standardization population.

[14]On page 8 of the manual, it is stated that item try-outs were conducted with deaf and mentally retarded children. (It is unclear whether, and if so to what extent, the normative population of deaf children included pupils who also were mentally retarded.)

[15]A concurrent validity study was conducted on hearing children constituting the non-handicapped portion of the normative sample. An investigation involving students at the Iowa School for the Deaf is cited as evidence for the test's concurrent validity with deaf children (see pages 13-14).

[16]On pages 20-21 of the manual three concurrent validity studies are reported, one of which included 21 mentally retarded children.

[17]On pages 19-20, several predictive validity investigations are reported. These studies involved subjects ranging in number from 9 to 28 who were not members of the standardization population.

[18]On pages 56-58 of the manual a literature review of approximately 100 studies is presented, an unspecified number of which employed "mentally retarded or institutionalized children."

[19]Additional concurrent validity studies are reported on pages 128-131 of the manual. It is unclear whether any of these investigations included handicapped individuals.

[20]On page 51 of the manual, the findings from an additional concurrent validity study are reported, which was conducted on 46 mentally retarded children.

[21]This number refers to the proportion of mentally retarded children included in the normative sample (see page 23 of the manual). It is unclear whether mildly and moderately handicapped pupils participated in the standardization.

[22]No indices of concurrent validity are reported for the standardization population. On page 63 of the manual, summarized results of other studies are presented, although no mention is made of handicapped children's participation.

[23]On page 3 of the Technical Report it is stated that the children "were drawn from classes indentified as average. Highly accelerated students, students with severe

reading disabilities, and students who were unable to function in the English language were excluded from the sample." This description of participants does not elimintate the possibility that mildly or moderately handicapped children participated in the norming process.

[24]On page 9-10 of the Technical Report, additional concurrent validity studies are reported, none of which appears to have included handicapped children.

[25]Several investigations of predictive validity are described on page 11 in the Technical Report. One study employed 62 "retarded readers."

[26]On page 57 in the manual it is claimed that the predictive validity of several subtests "has been studied extensively." However, there are no citations for these studies, no description of the criterion measures used, and no indications of the types of children included in the investigations.

[27]On pages 26-27 of the manual several additional concurrent validity studies are reported, none of which makes clear the possible participation of handicapped children.

[28]On page 25 of the manual it is reported that, "Pupils were eliminated who were known to have serious health or emotional problems. Those who stuttered or had other serious speech defects were not included. Any pupil who had been double-promoted or who had repeated a grade was considered unsuitable. Otherwise, random selection procedures were followed."

[29]"Whole classes were tested to obtain the normative data, except for samples of remedial reading cases totaling 300 pupils" (see page 14 of manual). From this passage it is unclear whether, and if so to what extent, handicapped children participated in this standardization effort.

[30]On page 43 of the manual it is stated that, among 36,527 pupils involved in the effort to calibrate the difficulty of the test, 1,224 were mentally retarded and in special classes. It is unclear whether additional mainstreamed handicapped children participated in the test's development.